

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 1, 2001, 16:16:37 ; Search time 91.75 Seconds
(without alignments)
15.839 Million cell updates/sec

Title: US-09-331-631a-3_COPY_29_73

Perfect score: 252
Sequence: 1 SEFDROEYBECKRQCMQLET.....RCVSOCDKREEDIDMSKYD 45

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 88757 seqs, 32294092 residues

Total number of hits satisfying chosen parameters: 88757

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 08
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_39:**

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	77.5	30.8	588	1	VCLB_GOSHI
2	72	28.6	605	1	VCLB_GOSHI
3	67.5	26.8	47	1	AGRP_IDEFCY
4	67	26.6	1170	1	TSP1_HUMAN
5	65	26.6	1170	1	TSP1_MOUSE
6	65	25.8	1173	1	TSP1_XENLA
7	63	25.0	524	1	SRP_SOYBN
8	57	22.6	724	1	ATIL_VACCV
9	57	22.6	726	1	ATIL_VACCV
10	57	22.6	1284	1	ATL_CAMPX
11	55	21.8	305	1	PPP6_RAT
12	55	21.8	795	1	ENPL_CHICK
13	54.5	21.6	244	1	YLS3_CAEEL
14	54.5	21.6	1021	1	YLS3_CAEEL
15	54	21.4	291	1	CGD1_BRARE
16	54	21.4	643	1	RO60_CAEEL
17	53.5	21.2	301	1	G197_HUMAN
18	53.5	21.2	1170	1	TSP2_BOVIN
19	53	21.0	702	1	ATIL_VAVR
20	52.5	20.8	91	1	UCRH_HUMAN
21	52.5	20.8	1172	1	TSP2_HUMAN
22	52	20.6	33	1	MBP1_MAIZE
23	52	20.6	37	1	CG23_LUPAN
24	52	20.6	82	1	C2_OXYNO
25	52	20.6	716	1	ENPL_RABIT
26	52	20.6	802	1	ENPL_MOUSE
27	52	20.6	803	1	ENPL_HUMAN
28	52	20.6	804	1	ENPL_CANFA
29	52	20.6	1188	1	TIR_ECOLI
30	51.5	20.4	106	1	COLA_HORSE
31	51.5	20.4	108	1	COLB_HORSE
32	51.5	20.4	147	1	LAMP_HABOF
33	51.5	20.4	298	1	HSL1_CANAL

34	51.5	20.4	525	1	Y107_MBTUA
35	51.5	20.4	563	1	MUC5_BOVIN
36	51.5	20.4	860	1	UL52_HSV6V
37	51.5	20.4	860	1	UL52_HSV6V
38	51.5	20.4	1557	1	DVA1_DICVT
39	51	20.2	418	1	CDI5_HUMAN
40	51	20.2	741	1	FIBA_CHICK
41	51	20.2	881	1	ARP8_YEAST
42	51	20.2	893	1	YM92_CAEEL
43	51	20.2	1231	1	CFAR_HUMAN
44	51	20.2	1370	1	Z261_HUMAN
45	50.5	20.0	264	1	RPOD_METTH

ALIGNMENTS

RESULT	ID	VCLB_GOSHI	STANDARD	PRT	588 AA.
AC	P09801				
DT	01-MAR-1989	(Rel. 10, Created)			
DT	01-MAR-1989	(Rel. 10, Last sequence update)			
DT	15-JUL-1999	(Rel. 38, Last annotation update)			
DE	VICILIN C72 PRECURSOR (ALPHA-GLOBULIN B).				
OS	Gossypium hirsutum (Upland cotton).				
OC	Eukaryota: Viridiplantae: Embryophyta: Tracheophyta: Spermatophyta: Magnoliophyta: eudicotyledons: core eudicots: Rosidae: eurosids II: Malvales: Malvaceae: Gossypium.				
RN	(1)				
RA	SEQUENCE FROM N.A.				
RA	Chian C.A., Pyle J.B., Legocki A.B., Dure L. III;				
RT	"Developmental biochemistry of cottonseed embryogenesis and germination. XVIII. cDNA and amino acid sequences of the members of the storage protein families."				
RL	Plant Mol. Biol. 7:475-489(1986).				
CC	-I- FUNCTION: SEED STORAGE PROTEIN.				
CC	-I- SUBCELLULAR LOCATION: CYTOLEDONARY MEMBRANE-BOUND VACUOLAR PROTEIN BODIES.				
CC	-I- SIMILARITY: TO OTHER 7S SEED STORAGE PROTEINS (PHASEOLIN, VICILIN, CONVICILIN, CONGLICININ, ETC.).				
CC	-----				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).				
CC	-----				
DR	EMBL: M16891; AAA33071.1; -				
DR	PIR: A30838; FMCNAB.				
DR	HSSP: P50477; ICAX.				
DR	INTERPRO: IPR001113; -				
DR	PFAM: PF00546; Seedstore_7s; 1.				
KW	Seed storage protein; Signal.				
FT	SIGNAL				
FT	CHAIN				
FT	SEQUENCE				
FT	588 AA: 69729 MM: 63E699B29AB8ADEB CRC64:				

Query Match 30.8%; Score 77.5; DB 1; Length 588;
Best Local Similarity 40.5%; Pred. No. 0.059; Length 588;
Matches 15; Conservative 9; Mismatches 12; Indels 1; Gaps 1;

QY 2 EFDRQYBECKRQCMQLETSGQMRRCVSOCDKREED 38
DB 81 EDPQRVYBECKRQCMQLETSGQMRRCVSOCDKREED 116

RESULT 2
VCLB_GOSHI STANDARD: PRT; 605 AA.
ID VCLB_GOSHI

```
AC P09799;
DT 01-MAR-1989 (Rel. 10, Created)
DT 01-MAR-1989 (Rel. 10, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE VICILIN GC72-A PRECURSOR (ALPHA-GLOBULIN A).
OS Gossypium hirsutum (Upland cotton).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
OC Malvales; Malvaceae; Gossypium.
RN [1]
RP SEQUENCE FROM N.A.
RA Chlan C.A., Borrito K., Kamalay J.A., Dure L., III:
RT "Developmental biochemistry of cottonseed embryogenesis and
RT germination. XIX. Sequences and genomic organization of the alpha
RT globulin (vicilin) genes of cottonseed.";
RL Plant Mol. Biol. 9:533-546(1987).
CC -1- FUNCTION: SEED STORAGE PROTEIN.
CC -1- SUBCELLULAR LOCATION: CYTOSOL; MEMBRANE-BOUND VACUOLAR PROTEIN
CC BODIES.
CC -1- SIMILARITY: TO OTHER 7S SEED STORAGE PROTEINS (PHASEOLIN, VICILIN,
CC CONVICILIN, CONGLYCININ, ETC.).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-slb.ch/announce/
CC or send an email to license@isb-slb.ch).
CC -----
DR EMBL: M19378; AAA33069.1; -
DR PIR: S06398; S06398.
DR HSSE: P50477; ICAX.
DR INTERPRO: IPR00113; -
DR PFM: PF00546; Seedstore_7s: 1.
KW Seed storage protein; Signal.
FT SIGNAL.
FT CHAIN.
FT SEQUENCE.
SQ SEQUENCE 605 AA; 71049 MW; C9DB9371C976953B CRC64;

Query Match 28.6%; Score 72; DB 1; Length 605;
Best Local Similarity 42.5%; Pred. No. 0.26;
Matches 17; Conservative 9; Mismatches 10; Indels 4; Gaps 3;

QY 1 SEED-ROEYECRQCMQLETSQGMRCVSQCCKRFEEDI 39
DB 33 SEDDPQQRREDCRRC-QLETRGRQEQ--DKCEDKSEIQL 69

RESULT 3
AGRP_LUFCY STANDARD: PRT; 47 AA.
ID AGRP_LUFCY
AC P56568;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE 6.5 KDA ARGININE/GLUTAMATE-RICH POLYPEPTIDE (6.5K-AGRP).
OS Luffa cylindrica (Smooth loofah) (Sponge gourd).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I;
OC Cucurbitales; Cucurbitaceae; Luffa.
RN [1]
RP SEQUENCE.
RP TISSUE=SEED;
RX MEDLINE=97357433; PubMed=9214759;
RX "Primary structure of 6.5K-arginine/glutamate-rich polypeptide from
RX the seeds of sponge gourd (Luffa cylindrica).";
RL Biosci. Biotechnol. Biochem. 61:984-988(1997).
CC -1- FUNCTION: STORAGE PROTEIN WHICH PROVIDES NITROGEN AND CARBON
CC RESERVES DURING GERMINATION AND SEEDLING GROWTH.
CC -1- MASS SPECTROMETRY: MW=5693.39; METHOD=MALDI.
```

```
CC -1- SIMILARITY: SOME, TO 7S SEED STORAGE PROTEINS.
KW Seed storage protein.
FT DISULFID 12 33
FT DISULFID 16 29
SQ SEQUENCE 47 AA; 5698 MW; 588B0EC82273AC05 CRC64;

Query Match 26.8%; Score 67.5; DB 1; Length 47;
Best Local Similarity 41.2%; Pred. No. 0.081;
Matches 14; Conservative 4; Mismatches 15; Indels 1; Gaps 1;

QY 5 ROEYECRQCMQLETSQGMRCVSQCCKRFEEDI 37
DB 6 RYEYCAVRQVAVHGVRQRCCQVCEKRLRE 39

RESULT 4
TSP1_HUMAN STANDARD: PRT; 1170 AA.
ID TSP1_HUMAN
AC P07996;
DT 01-AUG-1988 (Rel. 08, Created)
DT 01-AUG-1988 (Rel. 08, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE THROMBOSPONDIN 1 PRECURSOR.
GN THBS1 OR TSP1 OR TSP.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
RN [1]
RP SEQUENCE FROM N.A.
KC TISSUE-ENDOTHELIAL CELLS;
RX MEDLINE=87057617; PubMed=2430973;
RA Lawler J., Hynes R.O.;
RT "The structure of human thrombospondin, an adhesive glycoprotein with
RT multiple calcium-binding sites and homologues with several different
RT proteins.";
RL J. Cell Biol. 103:1635-1648(1986).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=89139590; PubMed=2918029;
RX Hennessy S.W., Frazier B.A., Kim D.D., Deckwerth T.L.;
RX Baumgartel D.M., Rotwein P., Frazier W.A.;
RT "Complete thrombospondin mRNA sequence includes potential regulatory
RT sites in the 3' untranslated region.";
RL J. Cell Biol. 108:729-736(1989).
RN [3]
RP SEQUENCE OF 1-397 FROM N.A.
RX MEDLINE=87157592; PubMed=3030396;
RX Kobayashi S., Eden-Mcutchan F., Framson P., Bornstein P.;
RT "Partial amino acid sequence of human thrombospondin as determined by
RT analysis of cDNA clones: homology to malarial circumsporozoite
RT proteins.";
RL Biochemistry 25:8418-8425(1986).
RN [4]
RP SEQUENCE OF 1-374 FROM N.A.
RX MEDLINE=86287276; PubMed=3461443;
RX Dixit V.M., Hennessy S.W., Grant G.A., Rotwein P., Frazier W.A.;
RT "Characterization of a cDNA encoding the heparin and collagen binding
RT domain of human thrombospondin.";
RL Proc. Natl. Acad. Sci. U.S.A. 83:5449-5453(1986).
RN [5]
RP SEQUENCE OF 1-166 FROM N.A.
RX MEDLINE=89291870; PubMed=2544587;
RX Laberty C.D., Glerman T.M., Dixit V.M.;
RT "Characterization of the promoter region of the human thrombospondin
RT gene. DNA sequences within the first intron increase transcription.";
RL J. Biol. Chem. 264:11222-11227(1989).
RN [6]
RP SEQUENCE OF 1028-1170 FROM N.A.
RX la Fleur M., Jobin C., Gauchier J., Kreis C.G.;
RL Submitted (XXX-1992) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: ADHESIVE GLYCOPROTEIN THAT MEDIATES CELL-TO-CELL AND
CC CELL-TO-MATRIX INTERACTIONS. CAN BIND TO FIBRINGEN, FIBRONECTIN,
```

CC LAMININ AND TYPE V COLLAGEN.
 CC -1- SUBUNIT: HOMOTRIMER, CROSS-LINKED BY DISULFIDE BONDS.
 CC -1- SIMILARITY: BELONGS TO THE THROMBOSPONDIN FAMILY.
 CC -1- SIMILARITY: CONTAINS 1 VMFC DOMAIN.
 CC -1- SIMILARITY: CONTAINS 3 EGF-LIKE DOMAINS.
 CC -1- SIMILARITY: CONTAINS 3 TYPE-1 TSP REPEATS AND 7 TYPE-3 TSP REPEATS (WHICH BIND CALCIUM).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: M25631; AAA36741.1; -
 CC EMBL: X04665; CAA28370.1; -
 CC EMBL: X14787; CAA32889.1; -
 CC EMBL: J04835; AAA61178.1; -
 CC EMBL: M99425; AAB59366.1; -
 CC PIR: A05172; A05172.
 CC PIR: A25812; A25812.
 CC PIR: A26155; A26155.
 CC PIR: A30140; A30140.
 CC PIR: A34274; A34274.
 CC HSSP: P35555; 1EMO.
 CC MIM: 188060; -
 CC INTERPRO: IPR000561; -
 CC INTERPRO: IPR000884; -
 CC INTERPRO: IPR001007; -
 CC PFAM: PF00008; EGF; 2.
 CC PFAM: PF00090; tsp_1; 3.
 CC PFAM: PF00093; vmc; 1.
 CC PROSITE: PS00022; EGF_1; FALSE_NEG.
 CC PROSITE: PS01186; EGF_2; 1.
 CC PROSITE: PS00092; TSP1; 3.
 CC PROSITE: PS01208; VMFC; 1.
 CC Glycoprotein: Cell adhesion; Calcium-binding; Heparin-binding; Repeat; EGF-like domain; Signal.
 CC SIGNL 1 18
 CC CHAIN 19 1170
 CC DOMAIN 19 232
 CC DOMAIN 316 373
 CC DOMAIN 379 548
 CC DOMAIN 549 690
 CC DOMAIN 723 950
 CC DOMAIN 951 1170
 CC REPEAT 379 430
 CC REPEAT 435 491
 CC REPEAT 492 548
 CC DOMAIN 549 587
 CC DOMAIN 588 645
 CC DOMAIN 646 690
 CC REPEAT 723 758
 CC REPEAT 759 781
 CC REPEAT 782 817
 CC REPEAT 817 840
 CC REPEAT 841 878
 CC REPEAT 879 914
 CC REPEAT 915 950
 CC SITE 926 928
 CC DISULFID 270 270
 CC DISULFID 274 274
 CC DISULFID 551 562
 CC DISULFID 556 572
 CC DISULFID 575 586
 CC DISULFID 592 608
 CC DISULFID 599 617
 CC DISULFID 620 644
 CC DISULFID 650 663
 CC DISULFID 657 676
 CC DISULFID 678 689

FT CARBOHYD 248 248 N-LINKED (GLCNAC. . .).
 FT CARBOHYD 360 360 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 708 708 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1067 1067 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CONFLICT 84 84 T -> A (TN REF. 2, 3 AND 4).
 FT CONFLICT 523 523 T -> A (TN REF. 2).
 SQ SEQUENCE 1170 AA: 129412 MW: 69B3EDEAE3A395E CRC64;
 Query Match 26.6%; Score 67; DB 1; Length 1170;
 Best Local Similarity 39.5%; Pred. No. 1.9;
 Matches 15; Conservative 6; Mismatches 11; Indels 6; Gaps 2;
 QY 13 ROCMOLE-----TSGOMRRC-VSQCCKRFEEDIDMSEKY 44
 DB 404 RSCDSLNRCGSSVOTRCHIOECCKRRKQDGSMSHW 441
 RESULT 5
 TSP1_MOUSE
 ID TSP1_MOUSE STANDARD; PRT; 1170 AA.
 AC P35441;
 DT 01-JUN-1994 (Rel. 29, Created)
 DT 01-JUN-1994 (Rel. 29, Last sequence update)
 DT 01-OCT-1996 (Rel. 34, Last annotation update)
 DE THROMBOSPONDIN 1 PRECURSOR.
 GN THBS1 OR TSP1.
 OS Mus musculus (mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE-92128941; PubMed-1774063;
 RA Lawler J., Duquette M., Ferro P., Copeland N.G., Gilbert D.J.,
 RT Jenkins N.A.;
 RL "Characterization of the murine thrombospondin gene.";
 RL Genomics 11:587-600(1991).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE-92147683; PubMed-1371115;
 RA Laherty C.D., O'Rourke K., Wolf F.W., Katz R., Seldin M.F.,
 RA Dixit V.M.;
 RL "Characterization of mouse thrombospondin 2 sequence and expression during cell growth and development.";
 RL J. Biol. Chem. 267:3274-3281(1992).
 RN [3]
 RP SEQUENCE OF 1-490 FROM N.A.
 RX MEDLINE-90375546; PubMed-2398070;
 RA Bornstein P., Alfi D., Devarayalu S., Framson P., Li P.;
 RL "Characterization of the mouse thrombospondin gene and evaluation of the role of the first intron in human gene expression.";
 RL J. Biol. Chem. 265:16691-16698(1990).
 CC -1- FUNCTION: ADHESIVE GLYCOPROTEIN THAT MEDIATES CELL-TO-CELL AND CELL-TO-MATRIX INTERACTIONS. CAN BIND TO FIBRINOGEN, FIBRONECTIN, LAMININ AND TYPE V COLLAGEN.
 CC -1- SUBUNIT: HOMOTRIMER, CROSS-LINKED BY DISULFIDE BONDS.
 CC -1- SIMILARITY: BELONGS TO THE THROMBOSPONDIN FAMILY.
 CC -1- SIMILARITY: CONTAINS 1 VMFC DOMAIN.
 CC -1- SIMILARITY: CONTAINS 3 EGF-LIKE DOMAINS.
 CC -1- SIMILARITY: CONTAINS 3 TYPE-1 TSP REPEATS AND 7 TYPE-3 TSP REPEATS (WHICH BIND CALCIUM).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: M62470; AAA50611.1; -
 CC EMBL: M62450; AAA50611.1; JOINED.
 CC EMBL: M62451; AAA50611.1; JOINED.

DR EMBL; M62452; AAA50611.1; JOINED.
 DR EMBL; M62453; AAA50611.1; JOINED.
 DR EMBL; M62454; AAA50611.1; JOINED.
 DR EMBL; M62455; AAA50611.1; JOINED.
 DR EMBL; M62456; AAA50611.1; JOINED.
 DR EMBL; M62457; AAA50611.1; JOINED.
 DR EMBL; M62458; AAA50611.1; JOINED.
 DR EMBL; M62459; AAA50611.1; JOINED.
 DR EMBL; M62460; AAA50611.1; JOINED.
 DR EMBL; M62461; AAA50611.1; JOINED.
 DR EMBL; M62462; AAA50611.1; JOINED.
 DR EMBL; M62463; AAA50611.1; JOINED.
 DR EMBL; M62464; AAA50611.1; JOINED.
 DR EMBL; M62465; AAA50611.1; JOINED.
 DR EMBL; M62466; AAA50611.1; JOINED.
 DR EMBL; M62467; AAA50611.1; JOINED.
 DR EMBL; M62468; AAA50611.1; JOINED.
 DR EMBL; M62469; AAA50611.1; JOINED.
 DR EMBL; M87276; AAA53063.1; JOINED.
 DR EMBL; J05606; AAA40431.1; JOINED.
 DR EMBL; J05605; AAA40431.1; JOINED.
 DR PIR; B42587; B42587.
 DR PIR; A37905; A37905.
 DR HSSP; P35555; 1EMO.
 DR MGD; MGI:9873; THBS1.
 DR INTERPRO; IPR000561; -.
 DR INTERPRO; IPR000884; -.
 DR INTERPRO; IPR001007; -.
 DR PFAM; PF00008; EGF_2.
 DR PFAM; PF00090; tsp_1; 3.
 DR PFAM; PF00093; wvc; 1.
 DR PROSITE; PS00022; EGF_1; FALSE_NEG.
 DR PROSITE; PS01186; EGF_2; 1.
 DR PROSITE; PS50092; TSP1; 3.
 DR PROSITE; PS01208; VMFC; 1.
 DR GLYCOPROTEIN; Cell adhesion; Calcium-binding; Heparin-binding; Repeat; EGF-like domain; Signal.
 FT SIGNAL 1 18
 FT CHAIN 19 1170
 FT DOMAIN 19 232
 FT DOMAIN 316 373
 FT DOMAIN 379 548
 FT DOMAIN 549 690
 FT DOMAIN 723 950
 FT DOMAIN 951 1170
 FT REPEAT 379 430
 FT REPEAT 435 491
 FT REPEAT 492 548
 FT REPEAT 549 587
 FT DOMAIN 588 645
 FT DOMAIN 646 690
 FT REPEAT 723 758
 FT REPEAT 759 781
 FT REPEAT 782 817
 FT REPEAT 818 840
 FT REPEAT 841 878
 FT REPEAT 879 914
 FT REPEAT 915 950
 FT SITE 926 928
 FT DISULFID 270 270
 FT DISULFID 274 274
 FT DISULFID 551 562
 FT DISULFID 556 572
 FT DISULFID 575 586
 FT DISULFID 592 608
 FT DISULFID 608 617
 FT DISULFID 617 644
 FT DISULFID 644 663
 FT DISULFID 650 663
 FT DISULFID 657 676
 FT CARBOHYD 678 689
 FT CARBOHYD 248 248
 FT CARBOHYD 360 360

FT CARBOHYD 708 708 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1067 1067 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1025 1025 F -> L (IN AAA53063).
 SQ SEQUENCE 1170 AA; 129646 MW; 0443E493615E7F06 CRC64;
 Query Match 26.6%; Score 67; DB 1; Length 1170;
 Best Local Similarity 39.5%; Pred. No. 1.9;
 Matches 15; Conservative 6; Mismatches 11; Indels 6; Gaps 2;
 QY 13 ROOMLE-----TSGCMRRC-VSQCDKRFREEDIDMSKY 44
 DB 404 RSCDSLNNRCESSVQTRCHIQECDKRFKDGSGMSHW 441
 RESULT 6
 TSP1_XENLA
 ID TSP1_XENLA STANDARD; PRT: 1173 AA.
 AC P33448;
 DT 01-JUN-1994 (Rel. 29, Created)
 DT 01-JUN-1994 (Rel. 29, Last sequence update)
 DT 01-OCT-1996 (Rel. 34, Last annotation update)
 DE THROMBOSPONDIN 1 PRECURSOR.
 GN THBS1 OR TSP1.
 OS Xenopus laevis (African clawed frog).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae;
 OC Xenopodinae; Xenopus.
 RN [1]
 RP SEQUENCE FROM N.A.
 RL Urry L.A., Ramos J., Duquette M., Desimone D.W., Lawler J.;
 Submitted (xx-1993) to the EMBL/Genbank/DBJ databases.
 CC -!- FUNCTION: ADHESIVE GLYCOPROTEIN THAT MEDIATES CELL-TO-CELL AND
 CELL-TO-MATRIX INTERACTIONS. CAN BIND TO FIBRINOGEN, FIBRONECTIN,
 LAMININ AND TYPE V COLLAGEN.
 CC -!- SUBUNIT: HOMOTRIMER, CROSS-LINKED BY DISULFIDE BONDS.
 CC -!- SIMILARITY: BELONGS TO THE THROMBOSPONDIN FAMILY.
 CC -!- SIMILARITY: CONTAINS 1 EGF-LIKE DOMAIN.
 CC -!- SIMILARITY: CONTAINS 3 EGF-LIKE DOMAINS.
 CC -!- SIMILARITY: CONTAINS 3 TYPE-1 TSP REPEATS AND 7 TYPE-3 TSP REPEATS
 (WHICH BIND CALCIUM).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 or send an email to license@sib-sib.ch).
 CC -----
 DR EMBL; L04278; -; NOT_ANNOTATED_CDS.
 DR HSSP; P00740; 11XA.
 DR INTERPRO; IPR000561; -.
 DR INTERPRO; IPR000884; -.
 DR INTERPRO; IPR001007; -.
 DR PFAM; PF00008; EGF_2.
 DR PFAM; PF00090; tsp_1; 3.
 DR PFAM; PF00093; wvc; 1.
 DR PROSITE; PS00022; EGF_1; FALSE_NEG.
 DR PROSITE; PS01186; EGF_2; 1.
 DR PROSITE; PS50092; TSP1; 3.
 DR PROSITE; PS01208; VMFC; 1.
 DR GLYCOPROTEIN; Cell adhesion; Calcium-binding; Heparin-binding; Repeat;
 EGF-like domain; Signal.
 FT SIGNAL 1 22
 FT CHAIN 23 1173
 FT DOMAIN 23 235
 FT DOMAIN 319 376
 FT DOMAIN 382 546
 FT DOMAIN 550 693
 FT DOMAIN 626 726
 FT DOMAIN 727 953
 FT DOMAIN 954 1173
 FT REPEAT 382 433

OY 12 KRCOMLETSQMRVCVSCDKREEDIMWSK 43
 DB 552 KRRVWEWELS-RLRRDIKECDK-YKEDLDKAK 581

RESULT 11

PP6_RAT ID PPP6_RAT STANDARD; PRT; 305 AA.
 AC 064620;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 15-JUL-1998 (Rel. 36, Last annotation update)
 DE SERINE/THREONINE PROTEIN PHOSPHATASE 6 (EC 3.1.3.16) (PP6) (PROTEIN DE PHOSPHATASE V) (PP-V).
 GN PPP6C OR PPV.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-SPRAGUE-DAWLEY; TISSUE=TESTIS, AND BRAIN;
 RX MEDLINE=94357899; PubMed=8077208;
 RA Becker W., Kentrup H., Klump S., Schultz J.E., Joost H.G.;
 RT "Molecular cloning of a protein serine/threonine phosphatase containing a putative regulatory tetrapeptide repeat domain.";
 RL J. Biol. Chem. 269:22586-22592(1994).
 CC -1- FUNCTION: MAY FUNCTION IN CELL CYCLE REGULATION.
 CC -1- CATALYTIC ACTIVITY: A PHOSHOPEPTIDE + H(2)O = A PEPTIDE + ORTHOPHOSPHATE (THIS ENZYME IS SERINE/THREONINE SPECIFIC).
 CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC (BY SIMILARITY).
 CC -1- TISSUE SPECIFICITY: HIGHEST LEVELS FOUND IN SPLEEN, BRAIN AND LUNG.
 CC -1- SIMILARITY: BELONGS TO THE PPP FAMILY OF PHOSPHATASES. PP-V SUBFAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@sib-sib.ch).
 CC -----
 CC EMBL: X77236; CAA54453.1; -
 CC DR HSSP: P08129; IFJM.
 CC DR INTERPRO: IPR000934; -
 CC DR PFAM: PF00149; STrphosphatase; 1.
 CC DR PRINTS: PR00114; STRPHPHASE.
 CC DR PROSITE: PS00125; SER_THR_PHOSPHATASE; 1.
 CC KW Hydrolyase; Iron; Manganese.
 CC FT METAL 53 53 IRON (BY SIMILARITY).
 CC FT METAL 55 55 IRON (BY SIMILARITY).
 CC FT METAL 81 81 IRON AND MANGANESE (BY SIMILARITY).
 CC FT METAL 113 113 MANGANESE (BY SIMILARITY).
 CC FT ACT_SITE 114 114 GENERAL ACID (BY SIMILARITY).
 CC FT METAL 163 163 MANGANESE (BY SIMILARITY).
 CC FT METAL 237 237 MANGANESE (BY SIMILARITY).
 CC SO SEQUENCE 305 AA; 35106 MW; A91F993FEBDF110 CRC64;

Query Match 21.8%; Score 55; DB 1; Length 305;
 Best Local Similarity 38.2%; Pred. No. 13;
 Matches 13; Conservative 2; Mismatches 19; Indels 0; Gaps 0;

OY 4 DROVEECKRQCMLETSQMRVCVSCDKRPEE 37
 DB 5 DLDKVEIARQCKYLPENDLRLCDYVCDLLEE 38

RESULT 12
 ENPL_CHICK STANDARD; PRT; 795 AA.
 ID ENPL_CHICK

AC P08110; Q90870; Q90869;
 DT 01-AUG-1988 (Rel. 08, Created)
 DT 01-AUG-1988 (Rel. 08, Last sequence update)
 DT 15-DEC-1998 (Rel. 37, Last annotation update)
 DE ENDOPLASMIN PRECURSOR (HEAT SHOCK 108 KDA PROTEIN) (HSP108) (HSP 108) (TRANSFERIN-BINDING PROTEIN).
 GN TRAI.
 OS Gallus gallus (Chicken).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae; Gallus.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=87076542; PubMed=3024703;
 RA Kuloma M.S., Weigel N.L., Kleinsek D.A., Beattie W.G., Connely O.M., March C., Zarucki-Schulz T., Schrader W.T., O'Malley B.W.;
 RT "Amino acid sequence of a chicken heat shock protein derived from the complementary DNA nucleotide sequence.";
 RL Biochemistry 25:6244-6251(1986).
 RN [2]
 RP SEQUENCE FROM N.A. AND PARTIAL SEQUENCE.
 RX MEDLINE=94220155; PubMed=8166742;
 RA Hayes G.R., Himmler B.S., Weiner K.X.B., Lucas J.J.;
 RT "A chicken transferrin binding protein is heat shock protein 108.";
 RL Biochem. Biophys. Res. Commun. 200:65-70(1994).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC TISSUE=OVIDUCT.
 RX MEDLINE=87117509; PubMed=3027654;
 RA Kleinsek D.A., Beattie W.G., Tsai M.J., O'Malley B.W.;
 RT "Molecular cloning of a steroid-regulated 108k heat shock protein gene from hen oviduct.";
 RL Nucleic Acids Res. 14:10053-10069(1986).
 RN [4]
 RP SEQUENCE FROM N.A.
 RC TISSUE=OVIDUCT;
 RA Forsgren M.;
 RT Submitted (XXX-1987) to the EMBL/Genbank/DBJ databases.
 CC -1- SUBUNIT: HOMODIMER; DISULFIDE-LINKED.
 CC -1- SUBCELLULAR LOCATION: ENDOPLASMIC RETICULUM LUMEN.
 CC -1- SIMILARITY: BELONGS TO THE HEAT SHOCK PROTEIN 90 FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@sib-sib.ch).
 CC -----
 CC EMBL: M14772; AAA48826.1; -
 CC DR EMBL: M31321; AAA48827.1; -
 CC DR EMBL: X04961; CAA28629.1; -
 CC DR PIR: A24461; HHCH08.
 CC DR PIR: JC2205; JC2205.
 CC DR HSSP: P02829; IAH8.
 CC DR INTERPRO: IPR000886; -
 CC DR INTERPRO: IPR001404; -
 CC DR PFAM: PF00183; HSP90; 1.
 CC DR PRINTS: PR00775; HEATSHOCK90.
 CC DR PROSITE: PS00014; ER_TARGET; 1.
 CC DR PROSITE: PS00298; HSP90; 1.
 CC KW Chaperone; Endoplasmic reticulum; Glycoprotein; Calcium-binding; Signal; Heat shock.
 CC FT SIGNAL 1 21 POTENTIAL.
 CC FT CHAIN 22 795 ENDOPLASMIN.
 CC FT CARBOHYD 61 61 N-LINKED (GLCNAC. . .) (POTENTIAL).
 CC FT CARBOHYD 106 106 N-LINKED (GLCNAC. . .) (POTENTIAL).
 CC FT CARBOHYD 216 216 N-LINKED (GLCNAC. . .) (POTENTIAL).
 CC FT CARBOHYD 444 444 N-LINKED (GLCNAC. . .) (POTENTIAL).
 CC FT CARBOHYD 480 480 N-LINKED (GLCNAC. . .) (POTENTIAL).
 CC FT CARBOHYD 501 501 N-LINKED (GLCNAC. . .) (POTENTIAL).
 CC FT SITE 792 795 PREVENT SECRETION FROM ER.

```

FT CONFLICT 206 206 D -> E (IN REF. 3 AND 4).
FT CONFLICT 267 267 V -> L (IN REF. 3).
FT CONFLICT 303 303 E -> Q (IN REF. 3).
FT CONFLICT 307 307 N -> D (IN REF. 3 AND 4).
FT CONFLICT 317 317 E -> H (IN REF. 3).
FT CONFLICT 378 378 C -> A (IN REF. 3).
FT CONFLICT 593 594 EG -> DR (IN REF. 3).
FT CONFLICT 653 653 W -> C (IN REF. 3).
FT CONFLICT 669 675 GMDISTN -> VFSS (IN REF. 4).
SQ SEQUENCE 795 AA: 91555 MW: BEIB29E1DBEC5A9A CRC64;

```

```

Query Match 21.8%; Score 55; DB 1; Length 795;
Best Local Similarity 23.8%; Pred. No. 12;
Matches 10; Conservative 12; Mismatches 20; Indels 0; Gaps 0;

```

```

OY 2 EFDROEYECRQCMQLETSQGMRCVSCDRCFEEDIDMSK 43
DB 581 EFDGKRFQVAKGVAFEESEKSKESREALEKEFEPLNMMK 622

```

RESULT 13

```

YL53_CAEEL STANDARD; PRT; 244 AA.

```

```

AC P34433;
DT 01-FEB-1994 (Rel. 28; Created)
DT 01-FEB-1994 (Rel. 28; Last sequence update)
DT 01-JUN-1994 (Rel. 29; Last annotation update)
DE HYPOTHEICAL 29.0 KDA PROTEIN F44E2.3 IN CHROMOSOME III.
CN F44E2.3.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Pelodetinae; Caenorhabditis.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RX MEDLINE=94150718; PubMed=7906398;
RA Wilson R., Alnscough R., Anderson K., Baynes C., Berks M.,
RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
RA Craton M., Dear S., Du Z., Durbin R., Favell A., Fraser A.,
RA Fulton L., Gardner A., Green P., Hawkins T., Hillier L., Jier M.,
RA Johnston L., Jones M., Kershaw J., Kirsten J., Laister N.,
RA Latreille P., Lightning J., Lloyd C., Mortimore B., O'Callaghan M.,
RA Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Showkhen R.,
RA Sims M., Smailton N., Smith A., Smith M., Sonhammer E., Staden R.,
RA Sulston J., Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K.,
RA Waterson R., Watson A., Weinstock L., Wilkinson-Spoat J.,
RA Wohldman P.;
RT *2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans.
RL Nature 368:32-38(1994).
CC -I- SIMILARITY: TO DNAJ.

```

```

CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----

```

```

CC EMBL: U23646; AAA28041.1; -.
DR PIR: S44822; S44822.
DR MORMPER: F44E2.3; CE00181.
KW Hypothetical protein.
FT DOMAIN 3 45 ARG/ASP/LYS-RICH.
FT DOMAIN 79 90 PRO-RICH.
SO SEQUENCE 244 AA: 28994 MW: CCE677FB01A2B18 CRC64;

```

```

Query Match 21.6%; Score 54.5; DB 1; Length 244;
Best Local Similarity 39.4%; Pred. No. 12;
Matches 13; Conservative 8; Mismatches 7; Indels 5; Gaps 2;

```

```

OY 9 EECRQCMQLETSQGMRCVSCD---KKREFD 38
DB 164 EBSRKKCKQLE--AELEKRVLEAESRKKFEED 194

```

RESULT 14

```

YL88_CAEEL STANDARD; PRT; 1021 AA.

```

```

AC P46582;
DT 01-NOV-1995 (Rel. 32; Created)
DT 01-NOV-1995 (Rel. 32; Last sequence update)
DT 01-NOV-1995 (Rel. 32; Last annotation update)
DE HYPOTHEICAL 111.9 KDA PROTEIN C34E10.8 IN CHROMOSOME III.
CN C34E10.8.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Pelodetinae; Caenorhabditis.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA Kirsten J.;
RL Submitted (JUN-1994) to the EMBL/Genbank/DBJ databases.

```

```

CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----

```

```

DR EMBL: U10402; AAA19070.1; -.
DR MORMPER: C34E10.8; CE01188.
KW Hypothetical protein.
SQ SEQUENCE 1021 AA: 111858 MW: F1B63DAA2632F5F CRC64;

```

```

Query Match 21.6%; Score 54.5; DB 1; Length 1021;
Best Local Similarity 22.0%; Pred. No. 46;
Matches 11; Conservative 15; Mismatches 13; Indels 11; Gaps 2;

```

```

OY 1 SEFDROEYECR-----QCMQLETSQGMRCVSCDRCFEEDIDMSK 43
DB 24 NEYDKRYGQCMVASYRTLVRCKQIRSEKELAKHGRC---EHVFSK 69

```

RESULT 15

```

CGD1_BRARE STANDARD; PRT; 291 AA.

```

```

AC Q90459;
DT 15-JUL-1998 (Rel. 36; Created)
DT 15-JUL-1998 (Rel. 36; Last sequence update)
DT 01-OCT-2000 (Rel. 40; Last annotation update)
DE G1/S-SPECIFIC CYCLIN D1.
CN CYCD1.
OS Brachydanio rerio (zebrafish) (zebra danio).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Ostariophysi;
OC Cypriniformes; Cyprinidae; Rasbortinae; Danio.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=96138542; PubMed=8547308;
RA Yarden A., Salomon D., Geiger B.;
RT "Zebrafish cyclin D1 is differentially expressed during early
RT embryogenesis.";
RT Biochim. Biophys. Acta 1264:257-260(1995).
CC -I- FUNCTION: ESSENTIAL FOR THE CONTROL OF THE CELL CYCLE AT THE G1/S
CC (START) TRANSITION.
CC -I- SUBUNIT: INTERACTS WITH THE CDK4 AND CDK6 PROTEIN KINASES TO FORM
CC A SERINE/THREONINE KINASE HOLOENZYME COMPLEX. THE CYCLIN SUBUNIT
CC IMPARTS SUBSTRATE SPECIFICITY TO THE COMPLEX.
CC -I- SIMILARITY: BELONGS TO THE CYCLIN FAMILY. CYCLIN D SUBFAMILY.

```


